pd.read\_csv (‘local file add’)

pd.read\_table (‘local file add’, sep = ‘,’)

|  |  |
| --- | --- |
| df.shape | df.info() |
| df.dtypes | df.describe() / .T *#* *to cut down the deci place* |
|  | df.columns.unique() |
|  | df.sort\_values (‘col1’, ascending = False) |
|  | df.sort\_values ([‘col1’, ‘col2’], ascending = [True, False]) |
|  | df.index.values *#* *prints out all the index values* |
|  | df.groupby (‘col1’) *# grpby is iterable. Can use for loop. Each element will be a tuple containing col1 as its first element and the subset of the original df for that col1 as it's second element* |
|  | df [‘col1’].mean () |
|  | df [‘col1’].value\_counts () *# can use iloc to extract the value as this is a Series data type* |
|  | df.rename (columns = {‘org header’: ‘new header’}, inplace = True) *# to rename column names* |
|  | df.[‘new col’] = df.[‘col1’] + df.[‘col2’] *# adding new col where new col = col1+col2* |
|  | df.drop([‘col10’, ‘col9’], axis = 1, inplace = True) *# axis = 1 refers to column. axis = 0 refers to row* |
|  | df.dropna () |
|  | df[‘col1’].fillna (value = ‘value1’, inplace =True) |
|  | df.groupby (‘col1’)[‘col2’].mean() |
|  | df.groupby (‘col1’)[‘col2’].agg ([‘count’, ‘mean’, ‘median’]) |
|  | df.groupby (‘col1’)[‘col2’].apply (lambda x: x.iloc[0,:]) |

*# To re-arrange columns in df:*

df = df[[‘col1’, ‘col2’, ‘col3’]]

*# To transpose list of list*

list2 = map (list, zip(\*list1)

df.loc [20:25, [‘col1’, ‘col2’]]

df.iloc [20:25, [0,1]]

# Pandas plot

df [‘col1’].hist()

graph = df [[‘col1’, ‘col2’]].plot (kind = ‘bar, stacked = True)

graph.set\_xticklabels (df.col1)

# Seaborn plot

sns.pairplot (df)

sns.distplot (df[‘column’], bins = 10, kde = False)